DOCUMENT RESUME

ED 450 071 SP 039 578

AUTHOR	Shaker, Paul
TITLE	Teacher Testing: A Symptom.
PUB DATE	2000-00-00
NOTE	25p.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Educational Policy; Elementary Secondary Education; Evaluation Methods; Knowledge Base for Teaching; Performance Based Assessment; Politics of Education; Public Education; Standardized Tests; State Standards; *Teacher Certification; *Teacher Evaluation; *Teacher Qualifications; Teaching Skills
IDENTIFIERS	Pragmatism; Teacher Knowledge

ABSTRACT

Current teacher testing is a symptom of what is wrong with American public education, rooted in invalid generalizations of method from one discipline to another. America's top educational policymakers are rarely educators, instead tending to be political leaders. The Interstate New Teacher Assessment and Support Consortium (INTASC) has produced a useful expression of model core standards for licensing teachers. Prototype classroom performance assessments and the INTASC Test of Teaching Knowledge are also being developed. Examining the 10 principles, and their subcategories of knowledge, dispositions, and performances, demonstrates how elusive teaching professional qualities are when attempting to reduce them to standardized testing. Nonetheless, political decisionmakers argue that standardized tests are what is available and economical today in teacher testing (making the assumption that questionable testing is better than no testing). Policymakers give little attention to the predictive validity of standardized instruments. The National Board for Professional Teaching Standards has adopted performance assessment, which is considered fair and accurate, as its methodology. Though pragmatic responses to social problems have characterized American democracy, they are missing in teacher education. Pragmatism competes with lesser dispositions associated with denial, prejudice, and false optimism in the realm of teacher supply and quality. (Contains 20 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Shaker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Teacher Testing: A Symptom

By Paul Shaker

ED 450 071

Teacher testing as it is currently construed is a symptom of what is wrong with American public education. The fervor for ever higher stakes associated with teacher testing illustrates how deeply entrenched our wrong-headed approach has become. With the inception of Title II of the Higher Education Act (Public Law 105-244), the federal government has begun a process of employing such tests not only for entry to the classroom by individuals, but for determining which institutions of higher education will have the right to prepare candidates for licensure. (This latter goal would be achieved by restricting the access of colleges and universities to federal education funding, such as student loans.) The evolution of standardized testing in American society has been documented extensively, most notably in None of the Above (Owen, 1985), and, more recently, in The Big Test (Lemann, 1999). Emerging from legitimate origins and egalitarian motives, the uses of these tests have moved into increasingly dangerous political waters. In America these punitive and unscientific applications are reaching their apotheosis today in high stakes K-12 student and teacher testing with consequences that affect the foundations of public education.

Origins

Thorndike is famously quoted as having claimed that everything of value exists in some quantity and can be measured. This attitude has, until recently, proven to be perfectly in tune with the spirit of our scientific and technological times. The revolutionary impact of science and technology in engineering, medicine, agriculture, and manufacturing in the past two hundred years has led many in Western society to conclude that the methods of positive, empirical science have controlling relevance to other fields of endeavor such as human service professions, psychology, politics and economics. Faith in standardized testing emerges from this expansive

SP039578

BEST COPY AVAILABLE

generalization about methods of inquiry. At a time when science itself has become more diverse on its frontiers (e.g., the Uncertainty Principle, chaos theory), political opinion and public policy can frequently be found advocating discredited theories from the past or making ineffective generalizations across disciplinary lines.

To put it another way, there is nothing postmodern about teacher testing. This endeavor is rooted in invalid generalizations of method from one discipline to another. It promises reductive certitude in a complex human service sphere. A hunger for simple answers has overwhelmed good judgment in the application of this approach to teacher evaluations. Authorities in educational research have over the past twenty years recognized a range of methods needed to effectively evaluate educational matters. One notable result has been the new eminence of qualitative research; a second is the performance assessment movement. *Although promising, scholarly evaluation movements have emerged in force within educational research, policymakers have shown little interest in keeping up with the cutting edge of inquiry in the field of education.*

Tribes

Generally we expect those with credentials and expertise to govern a field of endeavor or, at a minimum, to be intimately involved in policy decisionmaking in their field. The alternative was famously illustrated in Stalin's central bureaucratic control of the arts, social sciences, and science during the 1930s. Ideology reigned over open inquiry with disastrous results. Only during the past decade has the pope redressed the Vatican's interventions into astronomy, which culminated in the forced recantations of Galileo. Such abuses become comically obvious over the passage of time, but have had painful and destructive consequences when they took place.

In America the top policymakers for education are rarely educators or students of education. Instead, they tend to be political leaders, such as

Secretary of Education Richard Riley, or the governors of the states. When they consult on educational matters, they are likely to listen to a combination of "policy experts" (typically trained in variations of political science), business leaders, "hard" scientists, and a very few token, ideologically-inclined academicians (e.g., Chester Finn, Diane Ravitch, William Bennett, John Silber, Arthur Bestor, Alan Bloom, John Chubb, Terry Moe). Normally absent from these discussions are persons with deep experience in the human service professions, schools, or in colleges of education; respected, objective education scholars; or leaders of the stakeholder institutions and organizations in education. Having expertise or experience in schools, in educational organizations or in colleges of education appears to disqualify persons from being at or near the levers of power in our field. The ongoing evolution of teacher testing is more easily understood when we consider how little influence educators have in the development of policies that govern our profession and how much disdain our legitimate research is met with in the circles of power (Berliner, 1996).

In a contemporary society steeped in respect for expertise, formal preparation, advanced degrees, and peer review, why is the field of education governed through an alternative model of power distribution? Why are educators, numerous and omnipresent in society, by far the largest profession, so disenfranchised with respect to their own field of endeavor? A number of explanations are possible, but let us focus on one in particular here since it subsumes many of the others in its scope. In the popular mind, the exclusion of educators from authority in education may be seen as merited because "those who can do; those who can't teach." *That is, a popular bias exists against those who have forsaken free market competition for a career of tenure, government employment and 180-day work years. This bias is well entrenched but obscures more fundamental motives.* This type of invidious discrimination is not levied on clergy, social workers, philanthropy professionals and a host of others who work outside business and industry. Educators attract a special opprobrium.

A Native American scholar was asked, after a presentation on the horrors of her people's experience at the hands of the U. S. government, why it was that the persecution of Indians was so relentless and excessive, even after they posed no rational threat to white dominance. She responded that the tribes embodied an alternative ideology, a set of values at variance with those of Western Europe. Their social organization, less individualistic and materialistic, more oriented toward the group and common ownership of property, was perceived as a fundamental threat to the new dominant culture of North America and, as such, was forcefully eradicated with nearly genocidal consequences.

Is there a parallel with educators and their relation to the majoritarian view of Americans? Could the seeming irrationality of the governance structure of education and the perverse restrictions on educator input into decisionmaking in education be explained in this way? Educators are a large "tribe" in America, numbering in the millions. There are over three million teachers, for example, and approximately 50,000 professors of education, among half a million college and university faculty. The institution of education is monumental in size: occupying 25% of Americans on any given day as students, instructors, administrators, or support staff. Size makes us significant, but our values make us a threat. Educators represent an alternative path in American society. They do not by any means reject materialism, but they entertain competing priorities in their scheme of values.

Educators as a group embody an alternative approach to life in our society. For us values such as "the life of the mind," "aesthetic appreciation," and "human service" compete successfully with "consumerism," "property," "wealth," and "status." Within families and peer groups those who take this alternative path are often subject to pressure and criticism, accused of escaping reality (as "professional students") or of being failures since they earn fewer economic rewards. Educators are tolerated in society since there is a recognition that public education is necessary and, ultimately, only the

"educator-types" seem willing to stay in teaching and work with youth. This acceptance is tempered, however, by stringent controls on the extent to which educators control their own institutions lest this massive enterprise more overtly educate for a reconceptualized society, based more closely on the dispositions of the teachers themselves. There is a culture war at the heart of the teacher test debate and its roots run deep into American culture.

Validity and Culture

Another old saw of psychometrics goes like this: "What is intelligence? Intelligence is what intelligence tests measure." Through a century of evolution in the testing field, this tautology may be more accepted today than ever before. The types of timed, objective, standardized tests that have been spun from psychology's initial forays into IQ have demonstrated a strong positive correlation to one another and to the mysterious "G factor" we have come to identify as intelligence. Our consensus definition of intelligence has come to be defined as the ability to manipulate mathematical and verbal symbols in the decontextualized setting of multiple choice questions. This technique has been generalized to perform gatekeeper functions for selective colleges, scholarship programs, entry to professions and jobs, and a range of other highly significant opportunities in our society. Because this methodology has resonated with the dominant worldview of America, few challenges to it have prevailed and a comprehensive debate over its assumptions has not occurred.

The work of Howard Gardner in Multiple Intelligences (1993), Daniel Goleman in Emotional Intelligence (1995), and Robert Sternberg in Practical Intelligence (2000) are several clear examples among many of how the cutting edge of psychology has moved far beyond the concepts influencing policymakers in government who dictate the direction of education as well as the popular media who affirm an outdated understanding of human ability. Another cost of governing a profession with laymen is this type of resistance to new directions in the field. While researchers such as Gardner

and Sternberg have established that intelligence can be explained in a rich, multivariate fashion that better explains the diversified paths through which individuals succeed, those who aspire to teach continue, in at least 44 states (AFT, 2000), to be faced with archaic, linear examinations. Myers-Briggs Type Indicator data reaffirms the apparent unfairness of teacher tests through its positive correlation of high achievement on standardized tests with the "intuitive" preference and corresponding negative correlation of performance on these instruments with the "sensing" preference. The sensing preference is dominant among those who teach elementary school (Myers, 1998).

American minorities' complaints that timed, objective tests discriminate against them are regarded by many as unfounded whining and requests for special treatment. The courts typically hold that a adverse impact is acceptable if close examination of the test content supports the assertion that the information tested is essential for teachers (McDowell, 2000). In this way basic skills tests have established a legitimacy that professional practices and even subject matter content tests have not. More interesting, however, is ongoing research that suggests there is a deeper, more defensible case for the minority resistance to these tests. Here is how the New York Times reports the story:

> But the habits of thought—the strategies people adopted in processing information and making sense of the world around them—were, Western scholars assumed, the same for everyone, exemplified by, among other things, a devotion to logical reasoning, a penchant for categorization and an urge to understand situations and events in linear terms of cause and effect.
>
> ...
>
> In a series of studies comparing European Americans to East Asians, Dr. Richard Nisbett and his colleagues have found that

> people who grow up in different cultures do not just think about different things; they think differently.
>
> ...Easterners, the researchers find, appear to think more "holistically," paying greater attention to context and relationship, relying more on experience-based knowledge than abstract logic and showing more tolerance for contradiction. Westerners are more "analytic" in their thinking, tending to detach objects from their context, to avoid contradictions and to rely more heavily on formal logic (Goode, 2000).[1]

The America First argument at this point would be that minority cultures need to conform in their cognitive style, just as they need to conform to English as the lingua franca. This nativist argument is oblivious to a larger truth about an immigrant society such as the United States: our historic strength has been achieved through integrating new cultures with our own, not by imposition of a totalitarian conformity to the old, established ways. Particularly in the field of education of minorities, in which teachers are translators among cultures, it is self-defeating to create a teaching corps that is monochromatic in its cultural awareness and understanding. In addressing higher order cognition, the American tradition of standardized testing is increasingly called into question due to the biases of its approach. These biases are exercised toward individuals within the dominant culture and among cultural groups in a diverse society such as the U.S. Even mainstream behavioral science on its cutting edge is questioning the conventional testing assumptions.

In the context of teacher examinations, the error of most contemporary testing approaches is aggravated by the nature of teaching itself since among human endeavors none exceed teaching as a complex act. Not only is the teacher attempting to instruct in fundamentals while elevating the character of students, but also in a diverse society these goals are addressed

in a context of vast cultural and linguistic differences. The entire enterprise is undertaken in a large group setting, which assures that no one strategy can reach all students optimally. Instructional hours are limited; teachers' time and energy stretched in many, including non-academic, directions. Over 20% of American students are growing up in economic poverty, helping to render a large number wanting in what Maslow defined years ago as "deficiency needs." For example, such children are hungry, ill-clothed or housed, and/or lacking medical care. A teacher who is effective in such settings will demonstrate a wide range of interpersonal and communication skills, as well as command of subject matter. These skills may include a sense of theatre and performance; empathy; multicultural insight; an ability to employ irony, humor, and persuasion; means of demonstrating conviction and a moral stance; an ability to cope with significant stress; and so on. Outstanding teachers send a number of communications to different students during the same teaching moment. Such teachers are communicating various messages, both regarding content and affect, in a range of modalities to each of the students according to that student's needs and abilities. Needless to say, fine teaching is a triumphant expression of intelligence—broadly defined—as well as character.

The Interstate New Teacher Assessment and Support Consortium (INTASC) has produced one of the latest and best expressions of "model core standards" for licensing teachers. Prototype classroom performance assessments and the INTASC Test of Teaching Knowledge are also being developed. Examining the ten principles--and their subcategories of knowledge, dispositions, and performances--demonstrates how elusive teaching professional qualities are when one attempts to reduce them to standardized testing. Performance assessment offers much more promise in this regard, but the cost and fairness aspects even in this approach are not to be underestimated A sample principle and selected subcategories follows. Imagine the multiple choice questions that are valid to establish these abilities, particularly when we leave the subcategory of "knowledge":

Principle #5: The teacher uses an understanding of individual and group motivation and behavior to create a learning environment that encourages positive social interaction, active engagement in learning, and self-motivation.

Knowledge

...

The teacher knows how to help people work productively and cooperatively with each other in complex social settings.

The teacher understands the principles of effective classroom management and can use a range of strategies to promote positive relationships, cooperation, and purposeful learning in the classroom.

The teacher recognizes factors and situations that are likely to promote or diminish intrinsic motivation, and knows how to help students become self-motivated.

Dispositions

The teacher takes responsibility for establishing a positive climate in the classroom and participates in maintaining such a climate in the school as whole.

The teacher understands how participation supports commitment, and is committed to the expression and use of democratic values in the classroom.

The teacher values the role of students in promoting each other's learning and recognizes the importance of peer relationships in establishing a climate of learning.

The teacher recognizes the value of intrinsic motivation to students' life-long growth and learning.

The teacher is committed to the continuous development of individual students' abilities and considers how different motivational strategies are likely to encourage this development for each student.

Performances

The teacher creates a smoothly functioning learning community in which students assume responsibility for themselves and one another, participate in decisionmaking, work collaboratively and independently, and engage in purposeful learning activities.

The teacher engages students in individual and cooperative learning activities that help them develop the motivation to achieve, by, for example, relating lessons to students' personal interests, allowing students to have choices in their learning, and leading students to ask questions and pursue problems that are meaningful to them.

The teacher organizes, allocates, and manages the resources of time, space, activities, and attention to provide active and equitable engagement of students in productive tasks.

The teacher maximizes the amount of class time spent in learning by creating expectations and processes for communication and behavior along with a physical setting conducive to classroom goals.

The teacher helps the group to develop shared values and expectations for student interactions, academic discussions, and individual and group responsibility that create a positive

> classroom climate of openness, mutual respect, support, and inquiry.
>
> The teacher analyzes the classroom environment and makes decisions and adjustments to enhance social relationships, student motivation and engagement, and productive work.
>
> The teacher organizes, prepares students for, and monitors independent and group work that allows for full and varied participation of all individuals. (INTASC,1995)

In face of these realities about teaching, along come decisionmakers in government and political leaders who respond that all this may be true, but that which is available and economical today in teacher testing is standardized tests. Therefore, the argument goes, this approach is better than not screening at all. The poverty of this analysis is illustrated by a recent New York Times columnist's encounter with the New York Liberal Arts and Sciences Test for teacher licensure:

> ...my results made little sense. I got a perfect score, 300, on the science and math section—bizarre, considering that I failed precalculus in high school and never took another science course after tenth-grade chemistry.
>
> Meanwhile, I am mortified to report that my worst score—a 260—was on the essay.
>
> ...Although it is not perfect, the exam seems a necessary gatekeeping device for a profession whose purpose, after all, is imparting knowledge. (Goodnough, 2000)

So strong is the societal affection for such tests, that even faced with a "bizarre and mortifying" outcome, the reporter affirms the methodology.

But what about the assertion that questionable testing is better than no testing? These tests may not only be irrelevant and simplistic, but they may have a negative impact on teacher success in classrooms. At a minimum they serve to define the profession and teacher education curriculum in a wholly inadequate manner. In California, for example, the state legislature mandated a reading examination for elementary, and eventually special education candidates, which is derived from a set of over 140 objectives provided by the state. This test, The Reading Instruction Competency Assessment--RICA, is emblematic of the imposition of politics into the heart of the teacher education curriculum and the university regardless of the counsel of professional educators, their accrediting bodies, or California's own independent Commission for the Credentialing of Teachers. The purpose was to assure mastery of phonics methodology for the teaching of reading, which has of course become a politicized, overwrought initiative of America's rightists. RICA is much more focused and specific than the run of teacher tests, but is a clear example of how these tests are not without consequences in the preparation teachers receive.

In the initial round of testing, over 90% of candidates statewide passed California's RICA. Predictably this did not result in accolades from government or media. In approximately the same timeframe national publicity very much attended the failure of 53% of Massachusetts' candidates in the first administration of that state's new teacher licensure examination (Melnick & Pullin, 2000). What is missing in such testing events is the public and media understanding that passing scores and rates are often determined, not according to any research-based, criterion-referenced standard, but by political whim or a practical decision about how many licensed teachers the state needs to let through its gauntlet. Psychometricians may express their opinions of where the cut scores should fall, but those in political authority normally make the final judgments. These are not honest standards; they are floating standards, i.e., moving targets. The courts have for decades been involved in sorting out this issue

(as in *Groves v Alabama State Board of Education)* and support cut scores based on validation studies that meet current psychometric standards. These studies are themselves evidence, however, of the uncertainty that accompanies definitions of "good teaching" and the pedagogical knowledge such teaching presumes. Validation studies of this type function in their own world of assumptions—and those assumptions are not rooted in a clear link between the content validity of the tests and the knowledge or performance of successful teachers.

A legislator was speaking in Pennsylvania congratulating an audience of teacher educators on the pass rate of candidates on the state licensure examination (an NTE clone). Were we wise to inform him that the state department of education had made a policy decision to fail a certain percentage of test-takers and that the pass rate reflected nothing more than this internal decision?

Now that institutions are to be ranked and judged by their graduates' rate of test success, other misleading practices are being introduced. For example:

> ...While Title II requires only that institutions report test scores for their "program completers," the scores for *all* test takers—who may or may not be enrolled in the institution's teacher education program—sometimes find their way into local newspapers. When substantial numbers of these test takers fail to pass the state licensure exam, institutions' pass rates may be portrayed inaccurately, possibly with devastating headlines.
>
> Another concern for education schools is being assigned blame for poor candidate performance on subject matter exams, despite the fact that preparation in content areas generally occurs in other areas of the institution. (AACTE, 2000)

Institutions that serve minority candidates also suffer invidious comparisons in the testing contest due to the well-established difficulties such candidates have with such tests. This again brings to mind the assertion that a flawed evaluation is better than no evaluation and "does no harm." The initial results of the California RICA had whites passing at a rate of 94.2%, Latinos, 79.8%, Asians, 87.8%, and African-Americans, 77.3% (California Commission on Teacher Credentialing, 1999, p. 23). Recent California basic skills results (CBEST) show whites (non-Hispanic) passing at a rate of 81.2%, Mexican-American, 54.6%, Asian-Americans, 68.7%, and African-Americans, 42.4% (Schieffer, 2000, p. 2). The results for universities are not reported in the media with adjustments for the ethnic constitution of an institution's student cohort. So it follows that the universities who answer society's call to recruit and prepare minority teachers appear to nearly all who read the data to be the least effective preparers of teachers. At the same time elite universities and non-diverse private colleges gain credibility in teacher education based on the test-taking gifts and demographic advantages of their less heterogenous student bodies.

Along the way policymakers have given little attention to the predictive validity of standardized instruments, particularly in the field of teacher testing. In a parallel vein, for all of its breadth of adoption, the SAT is known to only marginally predict success during the first year of college—but do so no better than high school grades and rankings (Owen, 1985). NTE and similar teacher tests are presented by agencies such as ETS with no claims of correlation to candidate success in classrooms, i.e., predictive validity. ETS in fact cautions users not to claim predictive correlation between NTE and teacher performance. Testmakers claim that predicting success in classrooms is not their job. Rather, their task is to assure minimum professional knowledge among those licensed to teach. Identifying what this pedagogical knowledge might be--as has been argued here--is a daunting task and one that has not been convincingly done. The "panels of experts" upon whose judgment the content validity of the tests rests are employed to justify a simplification and reduction of the teaching

act that is heavily influenced by the disciplinary background of the test writers. There seems to be no interest on the part of policymakers to address these incongruities. Standardized tests are beyond legitimate scrutiny.

It is no accident that the National Board for Professional Teaching Standards, in an effort to establish credible evaluation, has adopted performance assessment as its methodology, and not standardized testing. The activities of NBPTS have given lie to claims that no alternative to conventional teacher tests was available. Performance assessment presents such a choice and accrediting bodies as well as some state licensure authorities are moving in this direction. Whether they can create cost-effective varieties of assessment remains to be seen. What may happen is that the assessment "label" will be placed on what are, fundamentally, standardized, objective teacher tests, so that a veneer of quality is laid on the traditional, inadequate methods. Should this occur the integrity of performance assessment will be corrupted and the term, like so many others in education, discredited by misuse. (Note the efforts of Reading Recovery® to protect its "brand name" through trademarking. We regularly see this exceptional program wrongly cited in justifying lesser, diluted versions of its approach. The consequences are predictable.)

Performance assessment has entered educational evaluation over a period of two decades and brought with it great promise of fairness and accuracy. Assessment can accommodate the complexity of the teaching act far better than unidimensional objective tests. In the case of the leading example of performance assessment, the results are not in on how truly useful and valid the NBPTS approach will prove to be. Up to this point the approximately thirty rubrics are not all in place for the range of teaching fields that are to be represented. There is the danger that prior expectations and the conventional views of assessors may influence the scoring of these performances; that presuppositions about what good teaching is (the phenomenon is described by Stigler and Heibert, 1999) may crowd out creative, nonconforming approaches.

> Cultural scripts are learned implicitly, through observation and participation, and not by deliberate study. ...
>
> People within a culture share a mental picture of what teaching is like. We call this mental picture a *script.* ...
>
> It is not hard to see where the scripts come from or why they are widely shared. ...All of us probably could enter a classroom tomorrow and act like a teacher, because we all share this cultural script (p. 86-87).

In any event, even a tyranny of performance conventions would put the debate on a much higher plane than now functions. Currently there is virtually no debate over the content of tests of professional knowledge of teachers since our society accepts that the questions are a kind of trivia crossed with shopworn assumptions about behaviorist educational psychology. Open, credible content analysis of teacher examinations does not take place, even in face of their significance. *This disconnect between the content of the tests and state and NCATE standards, teacher education curriculum, and independent guidelines, such as INTASC is testimony to the inadequacy of our standards setting-teacher education curriculum-evaluation continuum.*

American Pragmatism

The immediate origins of the current mania for teacher testing in America can be traced to the alarmist claims of A Nation at Risk (1983) in which the alleged failures of American education were likened to attacks by a foreign enemy. This poorly researched and argued document followed a cyclical tradition of criticisms of American education that traces back to at least the post-World War II era. The ongoing prejudice against educators, discussed earlier, provides a fertile medium for such periodic outbursts. The reaction

to Progressive Education in the form of the Life Adjustment Curriculum set off one cycle; Sputnik the next; the countercultural revolution of the Sixties another; falling SAT scores had their day; and economic doldrums and the rise of Japan's economy paved the way for A Nation at Risk. Subsequently and over time, however, America invited civic education and other life skills back into the curriculum; the nation eventually won the race to the moon uncontested; integrated the Sixties radicals (even to the level of President); reversed the imagined SAT slide (while playing by the slanted, non-psychometrically sound rules of the media debate); and left the Japanese far behind economically. Instead of celebrating the repeated triumphs of the scapegoat, public education, criticism continues unabated. In place of merited praise we witness a powerful privatization/voucher movement, punitive teacher testing (as in Massachusetts), a drop in teachers' real income in the 1990s, and continual vilification of teachers' unions even though they have joined the reform movement.

Among the most damaging of the attacks is a constant criticism of those who teach as inadequate for the challenge at hand. In contrast, how little recognition has been given to the repeated heroism of teachers at Columbine and the other sites of murderous school violence. Professions are founded on respect by society for the expertise and ethics of practitioners. Consequently, there are practical and demoralizing effects of this negative media drumbeat. Teacher testing in turn is based on the assumption by policymakers that unqualified persons have been taking over American classrooms and becoming fixtures there by earning unassailable tenure. In light of this "takeover," teacher tests are presented as a gatekeeper—or a means of "cleaning house"-- and assuring us of high quality teachers. This cause and effect argument is so distanced from reality, however, that it defies rational analysis. It is a truism in America that, come September, every classroom will have a "teacher" in the front of the room. In tens of thousands of cases, however, such persons will not be properly credentialed, professional teachers. They will either be college graduates who lack professional preparation, or teachers prepared and licensed in fields other

than the one to which they are assigned. In California the state's figures regularly show that ten percent of teachers lack licenses of any kind except "emergency"; and in high demand fields, such as mathematics, approximately 70% lack proper content preparation.[2]

Screening out candidates from the lower ranks of test-takers does not bring more and better candidates into our classrooms. Through perseverance these same persons typically find their way into teaching by back-door routes. The rewards of the profession continue to fall short in attracting and, more importantly, holding persons of the stature we all desire. Also it should be noted, unfair criticism, media bias, and a lack of popular respect further discourage strong candidates from entering or staying in the field. In America it is not possible to drive up teachers' salaries or improve their working conditions by creating a shortage of qualified professionals. We have a long tradition of filling vacant positions with notably unprepared persons. Nonetheless we continue to march down this fruitless path of bashing the profession in the media, throwing unfair screening hurdles before motivated candidates, leaving salaries depressed, and then filling positions with "any warm body."

Whatever became of American Pragmatism? When faced with societal challenges, the nation has often succeeded admirably by moving beyond ideology to find policies that worked in terms of their outcomes. Despite cries of socialism, the New Deal was propagated in the 1930s and at a minimum sustained national morale in the face of the Great Depression. Social Security and Medicare provided safety nets for the elderly by recognizing that millions of people, for whatever reason, would not or could not voluntarily prepare economically for old age. The Supreme Court, followed by Congress, transcended our legal traditions in civil rights, acknowledging that Jim Crow had to be abolished. In another instance, the Court reevaluated longstanding tradition and established a woman's right to choose with respect to pregnancy. Pragmatic responses to real social problems have characterized American democracy at its best. Teacher

testing; however, as a means of improving education, fits into a contrarian American pattern—one of denial of facts.

Denial of this type has failed historically when it is a source of public policy. For example, over a decade passed and organized crime boomed before America repealed its failed attempt at Prohibition in 1933. Currently we have a recrudescence of that type of failed "reform" in our address to the issue of substance abuse. As our rate of incarceration soars, along with associated social and real costs, movement toward the medicalization of substance abuse problems is stalled. We are similarly stymied in face of the problem of poverty and its relation to education. Policymakers and the media perpetuate the myth that schools alone can overcome all the environmental deficiencies facing a child and educate that child successfully. Hunger, abuse, constant transience have no impact on a child's learning in the eyes of pundits and for teachers to claim otherwise is to "hide behind excuses."

American Pragmatism appears irregularly in our public life and is capable of leading us to great social victories. This tradition competes, however, with lesser dispositions associated with denial, prejudice, and false optimism. In the realm of teacher supply and teacher quality we are not addressing the issue through our better instincts. By improving the compensation and working conditions for teachers we could attract and retain sufficient numbers of the candidates we desire. Mid-career teachers currently earn approximately $25,000 less than comparably prepared persons in other fields. Working conditions in many urban settings are hazardous—psychologically if not somatically. The toughest teaching assignments typically pay the least and are left to the newest teachers, as veterans move up the career ladder within a district, or to another district.

Ultimately there is no shortage of prepared teachers, hundreds of thousands have taken their credentials and talent and left or failed initially even to enter the profession. Society has not acquired or retained their talents for our

schools when in competition for it in the open marketplace of our economy. *In none of these policy areas do teacher tests make an impact: the tests are simply irrelevant to the great challenges facing public education. They are a distraction and a sideshow that take attention away from the reforms that need to be made.* If our profession were attracting and retaining in sufficient numbers the talent that it should, we would be moving away from testing as some professions have (e.g., architecture) because candidate competition for desirable jobs would play out successfully in the marketplace. The quality issues we have are not rooted in keeping candidates out of the profession, they are in society's failure to attract and retain the candidates we want in the profession.

Women are the most dramatic example of education's brain drain. Prior to the successes of the women's movement, approximately 5% of law and medical students were female. Thirty years later the rate of women's participation in these schools is close to 50%. This change is due overwhelmingly to the compensation and working conditions found in schools. The intrinsic attractiveness of teaching as a human endeavor is strong—our ability to attract fine professors gives evidence of this. But when the gap in salary and workplace ambience becomes too great, talent leaves teaching. The outflow of gifted women attests to this and is paralleled in the African-American community that has also provided teachers at a lessening rate. A similar analysis seems to fit both groups: teaching—the most accessible of professions to minorities and women—is increasing bypassed by talented members of those groups as equal opportunity has opened the doors to other professions with better working conditions and competitive salaries. Even the innate attractiveness of the act of teaching cannot capture for the profession the talent our society needs.

Conclusion

The American Educational Research Association has recently published a position statement on *High Stakes Testing in PreK-12 Education*

(AERA/APA/NCME, 2000) that is based on the *Standards for Educational and Psychological Testing, 1999* (AERA/APA/NCME, 1999) sponsored by AERA, the American Psychological Association and the National Council on Measurement in Education. These guidelines are also quite relevant to high stakes teacher testing and illustrate how far current practice has drifted from scientific and ethical responsibility. Tesing programs should have:

- Protection Against High-Stakes Decisions Based on a Single Test
- Adequate Resources and Opportunity to Learn
- Validation for Each Separate Intended Use
- Full disclosure of Likely Negative Consequences of High Stakes Testing Programs
- Alignment Between the Test and the Curriculum
- Validity of Passing Scores and Achievement Levels
- Opportunities for Meaningful Remediation for Examinees Who Fail High-Stakes Tests
- Appropriate Attention to Language Differences Among Examinees
- Appropriate Attention to Students with Disabilities
- Careful Adherence to Explicit Rules for Determining Which Students Are to be Tested
- Sufficient Reliability for Each Intended Use
- Ongoing Evaluation of Intended and Unintended Effects of High-Stakes Testing

As teacher tests are currently employed, it can be argued that none of these standards is adequately met and among the list all are truly critical to fairness and accuracy in testing. For example, the tests are often used as a solitary gatekeeping event. They are frequently used outside the purposes for which they were validated. Since teaching is so complex and little consensus for the knowledge base for the beginning teacher exists, there is inadequate alignment between the tests, state and national standards, and the

curriculum. The validity of passing scores is dubious. The unintended effects of this system—the demoralization and banning of certain candidates—has not truly been examined.

In some manner educators need to further the long march toward good practice in teacher testing. Enforcing high standards in the application of the tests is one overdue initiative, as is the propagation of performance assessment. The abuses demonstrated in this field are symptoms of the general misunderstanding and mistreatment educators experience in American society. In order to combat this disposition, teacher testing should be an immediate point of contact that is forcefully engaged between educators and our organizations and the larger political system. Our concept of ourselves as professionals, as well as the future character of education are at stake.

Footnotes

1. See also the work of Walter Ong, as in Orality and literacy: The technologizing of the work (new accents), (1988) New York: Routledge.

2. For California data, see www.ctc.ca.gov.

References

American Association of Colleges for Teacher Education. (August 14, 2000). Education schools nationwide faced with year of challenges. *Briefs.* P. 1.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (AERA/APA/NCME), Joint Committee on Educational and Psychological Testing. (1999). *Standards for educational and psychological testing 1999.* Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (AERA/APA/NCME), Joint Committee on Educational and Psychological Testing. (2000). *High stakes testing in preK-12 education.* Washington, DC: American Educational Research Association.

American Federation of Teachers. (2000). *Building a profession: Strengthening teacher preparation and induction.* Washington, DC: author.

Berliner, D., et al. (1996). *The manufactured crisis: myths, fraud, and the attack on America's public schools.* Los Angeles: Perseus Press.

California Commission on Teacher Credentialing. (1999). *Annual report on the Reading Instruction Competence Assessment (RICA) June 1998-June 1999 (Preliminary version).* Sacramento, CA: Author.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Goleman, D. (1995). *Emotional intelligence.* NY: Bantam.

Goode, E. (August 8, 2000). How culture molds habits of thought. *The New York Times,* p. D1.

Goodnough, A. (August 23, 2000). Proving fit to teach, by a half measure. *The New York Times,* p. A18.

Interstate New Teacher Assessment and Support Consortium. (1995). *INTASC core standards.* Washington, DC: The Council of Chief State School Officers.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy.* New York: Farrar Straus & Giroux.

McDowell, K. C. (Winter). Teacher competency tests: Disparate impact, disparate treatment, and judicial scrutiny. *The State Education Standard, 1(1),* pp. 45-47.

Melnick, S. & Pullin, D. (2000). Can you take dictation? Prescribing teacher quality through testing. *Journal of Teacher Education, 51(4),* pp. 262-275.

Myers, I. B., et al. (1998). *Myers-Briggs manual.* Palo Alto, CA: Consulting Psychologists Press.

National Commission on Excellence in Education. (1983). *A nation at risk.* Washington, DC: Author.

Owen, D. (1985). *None of the above: Behind the myth of scholastic aptitude.* New York: Houghton-Mifflin Co.

Schieffer, J. (2000). *CACTE comments.* Los Angeles: California Association of Colleges for Teacher Education.

Sternberg, R. J., et al. (2000). *Practical intelligence in everyday life.* Cambridge, UK: Cambridge University Press.

Stigler, J. W. & Hiebert, J. (1999) *The teaching gap.* New York: The Free Press.